Controlling Cluster Volume Sizes in Fuzzy c-Means Clustering

Sadaaki Miyamoto
Department of Risk Engineering
School of Systems and Information Engineering
University of Tsukuba
Ibaraki 305-8573, Japan
E-mail: miyamoto@risk.tsukuba.ac.jp

Naoya Kurosawa
Master's Program in Science and Engineering
University of Tsukuba
Ibaraki 305-8573, Japan
E-mail: s0323315@ipe.tsukuba.ac.jp

***Abstract*— The aim of this paper is to propose two methods of introducing a variable for controlling cluster volume sizes. The formulation is thus to define two new objective functions having the control variable. The reason why two different formulations are possible is described. The both methods have been motivated from referencing two different aspects of the entropy-based method of fuzzy c-means. Numerical examples show effectiveness of the proposed methods.**

I. Introduction

Fuzzy c-means clustering has been studied by many authors and applied to a variety of real problems. Nevertheless some fundamental problems remain unsolved yet. One of such problems is to control cluster volume sizes. The aim of this paper is to discuss this problem.

There are two types of objective functions of fuzzy c-means: one is the well-known function proposed by Dunn [1] and Bezdek [2] which we call here the standard method, while the other uses an additional term of entropy (e.g., [3]) which is called the entropy-based method. The latter has been extended by Ichihashi *et al*. [4] to incorporate the covariance matrix and the cluster volume size variables. The last method has moreover been proved to have a close relationship with the mixture normal distribution model [5]. On the other hand, the fuzzy covariance variable has been introduced to the standard method in a number of different ways [6], [7], but consideration of the cluster volume size is not extensively studied.

In this paper we consider two methods of introducing cluster volume size variables to the standard method, observing two features of the entropy-based method. As a result we have different objective functions corresponding to these features.

There are two main approaches to the fuzzy c-means clustering. One is the alternate optimization of an objective function; the other is to define the memberships, cluster centers, *etc* by an algorithm of fixed point iterations. The latter has great degree of freedom in choosing a part of calculation method and is *ad hoc*, and hence we stick to the first approach of alternate optimization, since establishing a new optimization method will induce new derivatives of *ad hoc* algorithms.

We show illustrative xamples and compare effectiveness of the proposed methods.

II. Fuzzy c-Means Algorithms

Below we omit discussion of covariance matrix [7] for simplicity, but the use of covariance variable to the methods herein is straightforward.

A. Basic formulations of fuzzy c-means

Let the objects for clustering be points in the p-dimensional Euclidean space: $x_k = (x_k^1, \dots, x_k^p)^t$ $(1 \le k \le n)$ and the clusters be denoted by C_i or simply by i $(1 \le i \le c)$. The membership matrix is $U = (u_{ik})$ $(c \times n)$ and $v_i = (v_i^1, \dots, v_i^p)$ are cluster centers. Moreover put $V = (v_1, \dots, v_c)$.

The basic alternate optimization algorithm of fuzzy c-means is the iteration of **FC1, FC2** and **FC3** as follows [2].

Basic Fuzzy c-Means Algorithm
FC0. Set the initial value of $\bar{V}$.
FC1. Solve $\min_{U \in M} J(U, \bar{V})$ and let $\bar{U}$ be the optimal solution.
FC2. Solve $\min_{V} J(\bar{U}, V)$ and let $\bar{V}$ be the optimal solution.
FC3. If the solution $(\bar{U}, \bar{V})$ is convergent, stop; else go to **FC1**.
End of FC.

where $M = \left\{ (u_{ik}) \,:\, u_{ik} \in [0,1], \sum_{i=1}^{c} u_{ik} = 1, \forall k \right\}$. As the objective function $J(U, V)$ the following two are discussed here.

$$J(U,V) = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^m \|x_k - v_i\|^2 \quad (m > 1) \tag{1}$$

$$J(U,V) = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik} \|x_k - v_i\|^2 + \lambda^{-1} \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik} \log u_{ik} \} \tag{2}$$

The former is a well-known function by Dunn [1] and Bezdek [2], which we will discuss later. On the other hand, the solutions for the latter entropy-based objective function are as follows.

$$u_{ik} = \frac{e^{-\lambda\|x_k-\bar{v}_i\|^2}}{\displaystyle\sum_{j=1}^{c} e^{-\lambda\|x_k-\bar{v}_j\|^2}}, \tag{3}$$

$$v_i = \frac{\displaystyle\sum_{k=1}^{n} \bar{u}_{ik} x_k}{\displaystyle\sum_{k=1}^{n} \bar{u}_{ik}} \tag{4}$$

B. Variable for controlling cluster volume size

For the latter method using the entropy term [3], a generalized objective function has been proposed [4], where an additional variable $\alpha = (\alpha_1, \cdots, \alpha_c)$ for controlling cluster volume sizes is used:

$$\begin{aligned} J(U, V, \alpha) &= \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}\|x_k - v_i\|^2 \\ &\quad + \lambda^{-1}\sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}\log \alpha_i^{-1} u_{ik} \end{aligned} \tag{5}$$

The constraint for α is

$$A = \left\{ \alpha \,:\, \sum_{i=1}^{c} \alpha_i = 1, \alpha_i \geq 0, i = 1, \ldots, c \right\}$$

Then the alternate optimization is as follows.

An Extended Algorithm of Fuzzy c-Means

FC'0. Set initial value of $\bar{V}$, $\bar{\alpha}$.

FC'1. Solve $\min_{U \in M} J(U, \bar{V}, \bar{\alpha})$ and let the optimal solution be $\bar{U}$.

FC'2. Solve $\min_{V} J(\bar{U}, V, \bar{\alpha})$ and let the optimal solution be $\bar{V}$.

FC'3. Solve $\min_{\alpha \in A} J(\bar{U}, \bar{V}, \alpha)$ and let the optimal solution be $\bar{\alpha}$.

FC'4. If the solution $(\bar{U}, \bar{V}, \bar{\alpha})$ is convergent, stop; else go to **FC'1**.

End of FC'.

The optimal solutions are shown below, where $\bar{u}_{ik}$, $\bar{v}_i$, $\bar{\alpha}_i$ are written as u_{ik}, v_i, α_i, respectively without confusion.

$$u_{ik} = \frac{\alpha_i e^{-\lambda\|x_k-v_i\|^2}}{\displaystyle\sum_{j=1}^{c} \alpha_j e^{-\lambda\|x_k-v_j\|^2}}, \tag{6}$$

$$v_i = \frac{\displaystyle\sum_{k=1}^{n} u_{ik} x_k}{\displaystyle\sum_{k=1}^{n} u_{ik}} \tag{7}$$

$$\alpha_i = \frac{\displaystyle\sum_{k=1}^{n} u_{ik}}{n} \tag{8}$$

III. Cluster Size Variables in the Standard Method

Let us introduce a variable for controlling cluster volume sizes into the standard method. We have two different approaches for this purpose.

A. First method

Notice that the solution U for (1) is

$$\begin{aligned} u_{ik} &= \left[\sum_{j=1}^{c} \left(\frac{\|x_k - v_i\|^2}{\|x_k - v_j\|^2} \right)^r \right]^{-1} \\ &= \frac{e^{-r\log\|x_k-v_i\|^2}}{\sum_{j=1}^{c} e^{-r\log\|x_k-v_j\|^2}} \end{aligned} \tag{9}$$

where $r = \frac{1}{m-1}$.

The first idea is to compare solutions of the standard and the entropy-based methods. Namely, what we wish is to alter the above solution to the next form

$$u_{ik} = \frac{\alpha_i e^{-r\log\|x_k-v_i\|^2}}{\displaystyle\sum_{j=1}^{c} \alpha_j e^{-r\log\|x_k-v_j\|^2}} = \frac{\alpha_i\|x_k - v_i\|^{-2r}}{\displaystyle\sum_{j=1}^{c} \alpha_j\|x_k - v_j\|^{-2r}}$$

We therefore employ the next objective function:

$$J(U, V, \alpha) = \sum_{i=1}^{c} \alpha_i \sum_{k=1}^{n} (\alpha_i^{-1} u_{ik})^m \|x_k - v_i\|^2 \tag{10}$$

For simplicity, put $d_{ik} = \|x_k - v_i\|^2$. Then the solutions of **FC'2**, **FC'3**, **FC'4** are as follows, where $r = \frac{1}{m-1}$.

$$u_{ik} = \left\{ \sum_{j=1}^{c} \left(\frac{\alpha_j}{\alpha_i} \right) \left(\frac{d_{ik}}{d_{jk}} \right)^r \right\}^{-1} \tag{11}$$

$$v_i = \frac{\displaystyle\sum_{k=1}^{n} (u_{ik})^m x_k}{\displaystyle\sum_{k=1}^{n} (u_{ik})^m} \tag{12}$$

$$\alpha_i = \left[\sum_{j=1}^{c} \left\{ \frac{\sum_{k=1}^{n} (u_{jk})^m d_{jk}}{\sum_{k=1}^{n} (u_{ik})^m d_{ik}} \right\}^m \right]^{-1} \tag{13}$$

B. Second method

Second method is based on a different idea. We observe that the solution (8) of the size variable for (2) is a simple average of the membership values for the cluster.

To find an objective function in which the alternate optimization derives this form of the solution is the second approach.

For this purpose, general values of the parameter m are unusable but we should employ the particular value of $m = 2$.

We propose the next function:

$$J(U,V,\alpha) = \sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ik})^2 d_{ik} - \lambda \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}\log\alpha_i \qquad (14)$$

with the same constraint $A = \{\, \alpha \,:\, \sum_{i=1}^{c}\alpha_i = 1, \alpha_i \geq 0,\, i = 1,\ldots,c \,\}$.

A simple application of the classical Lagrange multiplier solution to this objective function without considering the nonnegative property of u_{ik} and α_i leads to

$$u_{ik} = \frac{1 - \lambda \sum_{j=1}^{c} d_{jk}^{-1}\log\alpha_j}{d_{ik}\sum_{j=1}^{c} d_{jk}^{-1}} + \frac{\lambda\log\alpha_i}{d_{ik}} \qquad (15)$$

$$\alpha_i = \frac{\sum_{k=1}^{n} u_{ik}}{n} \qquad (16)$$

However, the nonnegative property sometimes fails in these equations and hence the solution requires more calculation in general.

Generally, there exists an index set $I(k)$ for an object x_k:

$$I(k) = \{\, j \,:\, u_{jk} > 0 \,\}$$

Then the solution for u_{ik} is

$$u_{ik} = \frac{1 - \lambda \sum_{j\in I(k)} d_{jk}^{-1}\log\alpha_j}{d_{ik}\sum_{j\in I(k)} d_{jk}^{-1}} + \frac{\lambda\log\alpha_i}{d_{ik}} \qquad (17)$$

instead of (15) (cf. [8]). Thus, how to find this index set is the problem to be solved. This problem is not easy in general and it is necessary to use an advanced optimization algorithm, but in the case of $c = 2$, the solution is greatly simplified:

I. Calculate u_{1k} and u_{2k} by (15).

II. If $u_{1k} < 0$ (resp. $u_{2k} < 0$), then put $u_{1k} = 0$ and $u_{2k} = 1$ (resp. $u_{2k} = 0$ and $u_{1k} = 1$).

IV. Numerical Examples

Three illustrative examples in Figures. 1–3 have been tested. For such large and small clusters, the standard method of fuzzy c-means will have misclassifications; a part of a large cluster will be judged to belong to a smaller cluster, which can be theoretically proved [9], since the classification rule induces the Voronoi regions.

Although the first and the second examples look similar, the second is more difficult to classify; the single link method [10] cannot separate the two clusters in the second example.

Table I compares the results of the standard fuzzy c-means and the two methods proposed herein. Ten trials with different initial values were tested for each of the three methods. The success in Table I means no misclassified points and the number of success out of the ten trials are shown in the table as *N. of succ*. The initial values were random selection of a hard partition of clusters.

The two methods here were applied to the three examples. For the first method, the results were successful without any misclassified objects (Figs. 4–6). The second method was successful for the first and the third examples, but there were misclassifications in the second example.

For the second method, we have used the simple algorithm at the end of the last section for the first two examples which has two clusters. In contrast, we have tested all possible $I(k)$:

$$I(k) \in \{\{1,2,3\},\{1,2\},\{1,3\},\{2,3\},\{1\},\{2\},\{3\}\}$$

and checked the resulting memberships are nonnegative. For such admissible memberships, we took the optimal solution of the minimum value of the objective function.

V. Conclusion

We have proposed two methods of introducing an additional variable for controlling cluster volume sizes into the standard method of fuzzy c-means, while such variable has already been used in the entropy-based method. The first approach uses the form of the memberships in the entropy-based method while the second approach employs the solution form of the size variable in the entropy-based method. In the both cases they are referring to the entropy-based method.

A problem in the second method is that sometimes the solution is sensitive to the parameter λ. Moreover the second method requires a more complicated calculation method when the number of clusters is more than two. An efficient algorithm has been developed for a similar problem [8], but no such algorithm for the second method herein has been found yet. Thus our recommendation at present is the first method, although there are rooms for further study of the both approaches. For example, an efficient and elegant algorithm for the second method is an interesting theoretical problem to be studied.

Acknowledgment: This study has partly been supported by the Grant-in-Aid for Scientific Research (B), Japan Society for the Promotion of Science, No.16300065.

References

[1] J.C. Dunn, A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, *J. of Cybernetics*, Vol.3, pp. 32–57, 1974.

[2] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum, New York, 1981.

[3] S. Miyamoto and M. Mukaidono, Fuzzy c - means as a regularization and maximum entropy approach, *Proc. of the 7th International Fuzzy Systems Association World Congress (IFSA'97)*, June 25-30, 1997, Prague, Czech, Vol.II, pp. 86–92, 1997.

[4] H. Ichihashi, K. Honda, N. Tani, Gaussian mixture PDF approximation and fuzzy c-means clustering with entropy regularization, *Proc. of the 4th Asian Fuzzy System Symposium*, May 31-June 3, 2000, Tsukuba, Japan, pp.217–221.

[5] R.A. Redner and H.F. Walker, Mixture densities, maximum likelihood and the EM algorithm, *SIAM Review*, Vol.26, No.2, pp. 195–239, 1984.

[6] E.E. Gustafson, W.C. Kessel, *Fuzzy clustering with a fuzzy covariance matrix*, IEEE CDC, San Diego, California, pp. 761–766, 1979.

[7] F. Höppner, F. Klawonn, R. Kruse, T. Runkler, *Fuzzy Cluster Analysis*, Wiley, Chichester, 1999.

[8] S. Miyamoto, D. Suizu, O. Takata, Methods of Fuzzy c-Means and Possibilistic Clustering Using a Quadratic Term, *Scientiae Mathematicae Japonicae*, Vol.9, pp. 17–33 (2003).

[9] S. Miyamoto, K. Umayahara, Methods in hard and fuzzy clustering, In: Z.Q.Liu and S.Miyamoto, eds., *Soft Computing and Human Centered Machines*, Springer, Tokyo, 2000, pp. 85–129.

[10] B.S. Everitt, *Cluster Analysis*, 3rd ed., Arnold, London, 1993.

TABLE I

NUMBER OF SUCCESSES OUT OF TEN TRIALS FOR THE THREE METHODS WITH THE USED PARAMETERS IN THE PARENTHESES

Method	*N. of succ.: Ex. 1*	*N. of succ.: Ex. 2*	*N. of succ.: Ex. 3*
Standard FCM	0	0	0
First Method	10 ($m = 1.2$)	10 ($m = 2.625$)	3 ($m = 1.2$)
Second Method	10 ($\lambda = 0.027$)	0	3 ($\lambda = 0.012$)

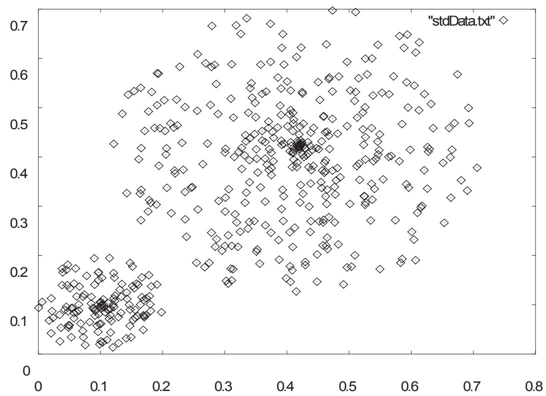

Fig. 1. An example of a large cluster and a small cluster

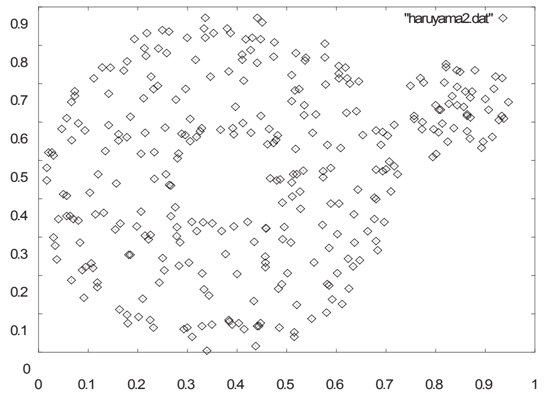

Fig. 2. Second example of a large cluster and a small cluster

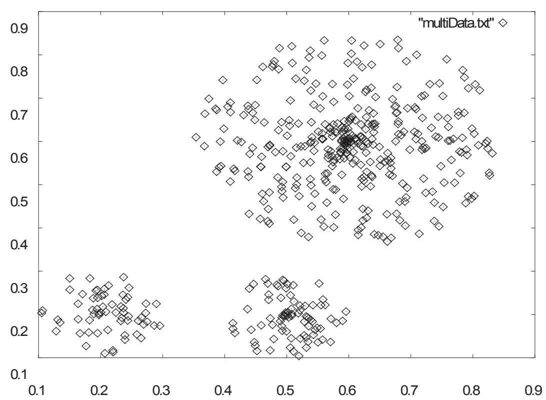

Fig. 3. Third example of three clusters

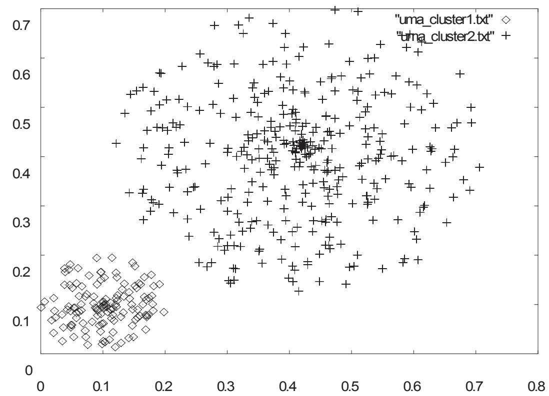

Fig. 4. A successful classification result for the first example

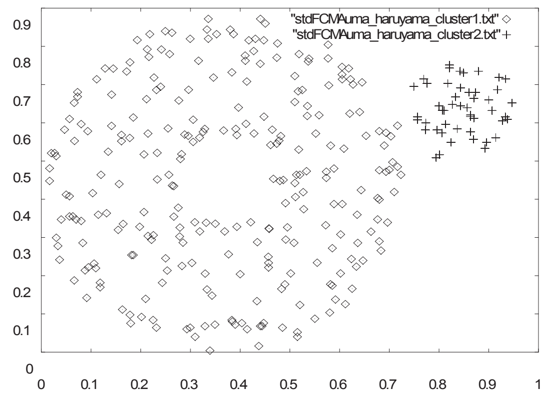

Fig. 5. A successful classification result for the second example

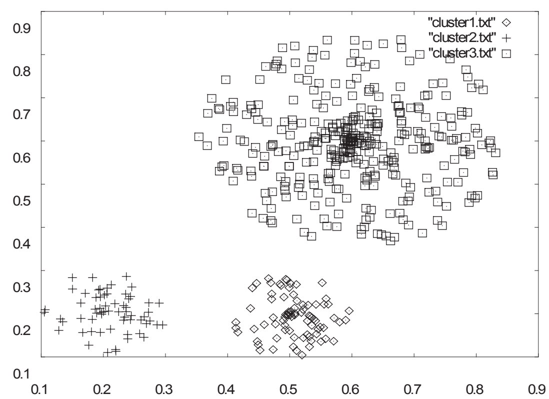

Fig. 6. A successful classification result for the third example